Modelling Conceptual Revolutions

PAUL THAGARD *University of Waterloo*

Like Clark Glymour, Don Ross does not appreciate what computer models can contribute to cognitive psychology and naturalistic philosophy of science.[1] This is not surprising, since the cognitive-computational approach differs dramatically from the conceptual systems and methodologies familiar to most philosophers. I will therefore begin by reviewing the use of computational models in cognitive science and then explain why the computational model ECHO (Explanatory Coherence by Harmany [*sic*] Optimization) is an essential part of my account of explanatory coherence and conceptual revolutions. This should make it obvious that my approach is not a search for anything like "computational foundations" as Ross suggests. I will also reply to some subsidiary objections and point to some recent work that expands the ideas in *Conceptual Revolutions.*

Many philosophers' view of reasoning divide it into two kinds, deductive and inductive, with first-order logic providing the standard way of thinking about deduction, and other formal accounts such as probability theory and confirmation theory providing the standard way of thinking about induction. In contrast, those of us studying scientific thought within cognitive science have a broader view of thinking, which emphasizes two components: structures and processes. In opposition to the standard philosophical assumption that sentence-like structures are the sole consituents of knowledge, in my book I described the importance of concepts and conceptual systems. Cognitive science also has a much richer view of processes than the traditional philosophical view of deduction and induction allow, making possible rigorous discussion of the nature of hypothesis formation, concept formation, analogy, problem solving and other important aspects of scientific thought. The interesting

Dialogue XXXV (1996), 155-59

question is: what would it take to shift those who have not seen the advantages of a cognitive approach away from the narrow logic-based view of scientific reasoning? Naturally, my answer is explanatory coherence: much more about the structure and development of science can be explained within the broader framework.

For investigation of mental processes, computer models are indispensable. Comprehension of their use requires noting the distinctions and the connections among four crucial aspects: theory, model, program and platform. A cognitive *theory* postulates a set of representational structures and a set of processes that operate on these structures. A computational *model* makes these structures and processes more precise by interpreting them by analogy with computer programs that consist of data structures and algorithms. Vague ideas about representations can be supplemented by precise computational ideas about data structures, and mental processes can be defined algorithmically. For the model to be tested, it must be implemented in a software program in a programming language such as LISP or C. This program may run on a variety of hardware platforms such as Sun Workstations or IBM™ personal computers (PCs), or it may be specially designed for a specific kind of hardware. Many kinds of structures and processes can be investigated in this way, from the rules and search strategies of some traditional sorts of artificial intelligence to the distributed representations and spreading activation processes of newer connectionist views.

The analogy between mind and computer is useful at all three stages of the development of cognitive theories: discovery, development and evaluation. Computational ideas about different kinds of programs often suggest new kinds of mental structures and processes. Developing a theory and a model often goes hand in hand with developing a program, since writing the program may lead to the invention of new kinds of data structures and algorithms that become part of the model and have analogs in the theory. Similarly, we often evaluate theory, model and program simultaneously, since one's confidence in the theory depends upon the validity of the model as shown by the performance of the program. The program can contribute to evaluation of the model and theory in three ways. First, it helps to show that the postulated representations and processes are computationally realizable. This is non-trivial, since many algorithms that seem at first glance reasonable are intractable and could not be applied to large problems on real computers.[2] Second, in order to show not only the computational realizability of a theory but also its psychological plausibility, the program can be applied qualitatively to various examples of thinking. Third, to show a much more detailed fit between the theory and human thinking, the program can be used quantitatively to generate detailed predictions about human thinking that can be compared with the results of psychological experiments. Cognitive theories by themselves are normally not precise enough to generate such quantitative predictions, but a model and program may fill the gap between theory and observation.

Now it is possible to answer Ross's question about the purpose of the computational model ECHO. He is puzzled about whether it adds anything to the theory of explanatory coherence (TEC), which I used to account for belief change in the major scientific revolutions. As suggested in the last paragraph, ECHO is crucial to the evaluation of TEC in three ways. First, the general problem of choosing the best explanation of a set of observations is known to be computationally intractable (Bylander, Allemang, Tanner and Josephson 1991). Any practical account of theory choice needs a computational implementation to demonstrate that it is realizable. For example, a crude Bayesian theory that requires a full joint distribution of probabilities is not realizable for large examples, since the provision of 2^n conditional probabilities for n propositions will quickly exhaust the capacity of any physical system.[3] ECHO has the desirable property that the number of cycles of activation updating required to determine the best explanation does not increase with the number of propositions to be evaluated (Thagard, forthcoming).

The second way in which ECHO is crucial to the evaluation of TEC is that it made possible detailed and effective application to numerous cases in the history of science. The largest of these, Newton versus Descartes and Copernicus versus Ptolemy, each involved more than 100 propositions each and are, I believe, the most detailed analyses of cases of scientific reasoning ever done (Nowak and Thagard 1992a, 1992b). ECHO shows that TEC can be applied in great detail to real historical cases, directly based on the writings of the relevant scientists. Contrary to Ross's assertion that computer simulation leads to gross simplification, this methodology encourages much more serious analysis of cases than is normally done in historical philosophy of science.

Finally, the third way in which ECHO has served to evaluate TEC is in the quantitative fit to psychological experiments. The theory and model have led to a series of psychological experiments designed to test it as an account of human thinking involving competing hypotheses. Using ECHO to generate predictions about human reasoning, psychologists have designed experiments that have shown a good fit between how the theory of explanatory coherence describes reasoning and how people actually do reason, in the social domain and in examples relevant to science education (Miller and Read 1991; Read and Marcus-Newhall 1993; Read and Miller 1993; Schank and Ranney 1991, 1992). These studies provide evidence that inference to the best explanation based on explanatory coherence is part of the cognitive apparatus of people in general, as well as of the scientists who use it more explicitly. TEC and ECHO are intended to be models of how scientists actually think, as well as of how they should think.

Ross erroneously implies that ECHO and computational philosophy of science in general depend on a kind of computational internalism that ignores physical and social environments. An expanded version of *Conceptual Revolutions* would indeed deal more extensively with the physical environments that had a major effect on the observations and experiments of scientists such as Lavoisier and Darwin, and it would also describe more fully their social environments. But these additions would have to mesh with the account of the conceptual changes that took place in individual scientists. Much of my work in the past couple of years has been concerned with overcoming the acknowledged omission of social factors in the discussion of scientific revolutions. Cognitive and social explanations of the development of science are best seen as complementary rather than competitive. What is needed is an account of how cognitive and social accounts of science can be integrated to provide unified naturalistic explanations of a wide variety of features of scientific development. I recently proposed that concepts from distributed artificial intelligence can be used to tie together cognitive and social accounts of science (Thagard 1993b), and have classified the explanation schemas used by cognitive and social theorists in order to show how unified schemas can be produced and used (Thagard 1994). I look forward to the expansion of current cognitive models of science and their further integration with social models. But taking physical and social environments into account supplements rather than replaces the kind of cognitive-computational explanation of scientific thinking that I and others have been offering.

In sum, Ross has seriously underestimated what computational ideas and models can contribute to naturalistic philosophy of science.

Notes

1 This note is a response to Don Ross's critical notice of my book, *Conceptual Revolutions.* My intervention incorporates some responses I made to comments by Clark Glymour and Ron Giere at the 1993 Eastern Division meeting of the American Philosophical Association in a symposium on the book. A few paragraphs are drawn from the English text that was translated for the Preface to the Italian edition of the book published by Guerini e Associati. This research is supported by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada.

2 For a philosophical discussion of intractability, see Thagard 1993b.

3 Much more sophisticated Bayesian approaches have been developed. See, for example, Pearl 1988, and see Thagard (forthcoming) for a comparison of explanatory coherence theory with Bayesian networks.

References

Bylander, T., D. Allemang, M. Tanner and J. Josephson
1991 "The Computational Complexity of Abduction." *Artificial Intelligence*, 49: 25-60.

Miller, L., and S. Read

1991 "On the Coherence of Mental Models of Persons and Relationships." In *Cognition in Close Relationships.* Edited by F. Fincham and G. Fletcher. Hillsdale, NJ: Erlbaum, pp. 69-99.

Nowak, G., and P. Thagard

1992a "Copernicus, Ptolemy, and Explanatory Coherence." In *Cognitive Models of Science.* Edited by R. Giere. Minneapolis: University of Minnesota Press, pp. 274-309.

1992b "Newton, Descartes, and Explanatory Coherence." In *Philosophy of Science, Cognitive Psychology and Educational Theory and Practice.* Edited by R. Duschl and H. R. Hamilton. Albany, NY: SUNY Press, pp. 69-115.

Pearl, J.

1988 *Probabilistic Reasoning in Intelligent Systems.* San Mateo, CA: Morgan Kaufman.

Read, S., and A. Marcus-Newhall

1993 "The Role of Explanatory Coherence in the Construction of Social Explanations." *Journal of Personality and Social Psychology,* 65: 429-47.

Read, S. J., and L. C. Miller

1993 "Explanatory Coherence in the Construction of Mental Models of Others." In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum, pp. 836-41.

Schank, P., and M. Ranney

1991 "Modeling an Experimental Study of Explanatory Coherence." In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum, pp. 892-97.

1992 "Assessing Explanatory Coherence: A New Method for Integrating Verbal Data with Models of On-line Belief Revision." In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum, pp. 599-604.

Thagard, P.

1992 *Conceptual Revolutions.* Princeton, NJ: Princeton University Press.

1993a "Computational Tractability and Conceptual Coherence: Why Do Computer Scientists believe that P ≠ NP?" *Canadian Journal of Philosophy,* 23: 349-64.

1993b "Societies of Minds: Science as Distributed Computing." *Studies in History and Philosophy of Science,* 24: 49-67.

1994 "Mind, Society, and the Growth of Knowledge." *Philosophy of Science,* 61, 4: 629-45.

Forthcoming "Probabilistic Networks and Explanatory Coherence. In *Automated Abduction: Inference to the Best Explanation.* Edited by P. O'Rorke and J. Josephson. Menlo Park, CA: AAAI Press.